

Credit Card Business in Malaysia: A Data Analytics Approach

Mohamed Khaled Yaseen¹

MSc Data Science, School of Computing
Asia Pacific University of Technology & Innovation
Kuala Lumpur, Malaysia

Mafas Raheem², V. Sivakumar³

Academic, School of Computing
Asia Pacific University of Technology & Innovation
Kuala Lumpur, Malaysia

Abstract—The revolution of big data has made resonance in the banking sector especially in dealing with the massive amount of data. The banks have the opportunity to know about the customer's opinions and satisfaction regarding their products by analyzing the data gathered every day. So, the banks can transform these data into high-quality information that allow banks to improve their business especially in credit cards which is becoming a short-term business for the banks nowadays. Further, the sentiment analysis has become immense in the field of data analytics especially the customers' opinion makes a huge impact in making profitable business decisions. The outcome of the sentiment analysis does assist the banks to know the deficiencies of their product and allow them to improve their products to satisfy the customers. From the sentiment analysis, 45% of the customers were negative, 30% were positive and 25% were neutral towards the credit card facility offered by the commercial banks. Also, the prediction of credit card customer satisfaction will contribute in a significant way to create new opportunities for the banks to enhance their promotion aspects as well as the credit card business in future. Random Forest algorithm was applied with three various experiments utilizing the normal data, balanced data and the optimized model with the normal data. The optimized model with the normal data obtained the highest accuracy of 87.38% followed by the normal dataset by 85.82% and the least accuracy was for the balanced dataset by 82.83%.

Keywords—Credit card; predictive analytics; random forest; sentiment analysis; banking

I. INTRODUCTION

The credit card or "plastic money" is a digital money lending service offered by various financial firms which include banks and any other financial institutions. The main aim of this card is to allow its users to borrow money at the sales point to allow them to complete their purchases easily and conveniently. However, this convenience may result in problems for the users as the users should repay the banks the amount they borrowed along with the specified interest rates at the end [1].

Based on the study conducted on marketing tools, the banks try to acquire more customers to purchase credit card services [2]. The findings showed that credit card facility will attract more customers when the banks offer more incentives. This states that the incentives are an effective marketing supportive tool in convincing the customers to use a particular card in their specific purchasing aspects. Incentives such as

cash rebates, lower interest, and airline miles may drive customers to spend more without thinking about the future [3]. Besides, another study stated that if the credit cards users are not fully aware of the incentives and the impact of their usage pattern of the credit cards might increase their future debt.

In the past few years, credit cards usage has been increased due to the incentives offered by financial institutions. At present, many credit cards users don't fully understand how those cards are operating. Yet, they are still using it just because of the offers they get from the cards. Nowadays people keep buying and exceeding their limits without recognizing that they are putting themselves in serious debt. However, if the credit cards are managed properly and users are fully aware of the pros and cons, it will be very beneficial for them in terms of its ease of use and convenience as it is the best alternative of cash [4].

In Malaysia, there are 3.6 million credit card users as of June 2017 and it was recorded that 18% of Malaysians between the age of 20 and 74 are using the credit cards based on the statistics provided by the Malaysian Department of statistics in 2016 [5]. In the meantime, the credit card outstanding balance was 36.9 billion Malaysian ringgits, where the overdue balance was 2.3 billion Malaysian ringgits which represent 7.3% [5]. It was also noted from the same source, the residual balance which has not exceeded 3 months of the repayment date was 2.3 billion Malaysian ringgits representing 6.2%, whereas the balance which has already exceeded 3 months was 0.4 billion Malaysian ringgits representing 1.1%. Furthermore, the statistics proved that in general, 43.6% of credit card users have paid their debts as well as another group of credit card users representing 43.6% settled leastwise 5% of their debts, whereas 12.8% of credit card users did not settle their debts. Besides, the aspect of insolvent cases, nearly 845 credit card users whose age is under 30 has declared their bankruptcy by the first half of 2017 [5].

This article is based on the study of credit card business in Malaysia and presented thematically in the following sections. Section 2 discusses literature regarding the relevant techniques and Section 3 covers the methodology followed in this study. Further, Sections 4 and 5 elaborates the descriptive and predictive analytics respectively which were conducted to offer actionable insights to the practitioners of the concerned domain.

II. LITERATURE REVIEW

A. Credit Card Business in Malaysia

The increase of the credit card users caused indiscriminate spending by the customers, which triggered various side effects as well. Numerous Malaysians happen to show excessive purchasing behaviour which is also known as "compulsive purchasing" [6]. Furthermore, it is considered easier for Malaysians to get a credit card rather than obtaining a loan facility due to the various complicated procedures. As a result of that, the millennials have adapted to move on with the modern lifestyle with the support of the credit cards. On the other hand, most of the credit card users have debts to settle which was due to the extensive use of the credit card facilities.

B. Bank Customers and Credit Cards

In general, consumers are inclined to recklessly using credit cards sometimes which ends up by overspending and having large amounts of debts. A study was conducted by interviewing fifteen young customers to grasp their perceptions and considerations regarding the use of credit cards. The results showed that there is a switch in spending behaviour. It was shown that the mind-set of the older generation is to save and spend later, whereas the mind-set of the younger generation is to borrow and spend now and pay later. The awareness of the advantages and disadvantages of the credit card facilities was also investigated in the same study. The results showed that young consumers are aware of the disadvantages which can cause a problem for them in the future, but they still use it and sometimes overspend due to the bonuses, points, discounts, gifts, and so on. Also, the world is moving toward a cashless society and the young customers are feeling satisfied with these cards as it offers them safety instead of carrying cash as well it offers them ease of use and convenience [7].

C. Factors Affecting Credit Card Usage

In Malaysia, numerous researchers have been examining the factors affecting the usage of credit cards which are influencing the customer's spending patterns. For instance, it was stated from the study conducted in Malaysia, that the age has a considerable influence on credit card usage and their behaviour. Moreover, the level of income and material status has a significant effect on credit card usage where customers who have high income and married are likely to spend more using their credit cards compared to those who have low income and single. Also, this study considered the bank policies in which the findings showed that this factor has a significant influence where customers are likely to pay more with credit cards when they are offered with some benefits by the banks such as longer period and low annual interest rates [8].

Numerous studies have been done in many countries examining the usage of credit card and the customer's behaviour. For instance, India as one of the biggest markets in the world, researchers have conducted studies on several factors such as gender, age, convenience, and sense of fulfilment. Results showed that convenience was the most influential where the ease of the use helps in increasing credit

card usage. Also, based on the age factor, the young customers spend more with their credit cards compared to the aged customers who still prefer to use cash, and this had come into agreement with the studies conducted in Malaysia as well. Furthermore, the gender factor in India is influencing differently compared to the Malaysian context where in India the males are likely to spend more through their credit cards than females. The sense of fulfilment factor also has influenced the usage of credit card among Indians where individuals who have credit cards feel that it is an achievement in their lifestyle [9].

Similarly, a survey in Klang Valley area in Malaysia was conducted to examine the socio-economic and demographic factors including level of income, age, gender, and occupation towards the usage of the credit cards [10]. The results stated that gender does not influence the credit card usage, where the personal background of the individual has a considerable influence on usage. Also, this study mentioned that the credit card will be used for spending in various situations where some use it on basic things and others use it for retail purchases and proved that the customer profile has a significant relationship with the credit card usage [10] [11].

D. Sentiment Analysis in the Credit Card Business

Nowadays, managing the customers' feedback has turned into a complicated activity due to the incorporation of various sources such as surveys on the customer's satisfaction, customers review and social media feedbacks. However, the capability to quickly grasp the feeling of the customers about the products is considered valuable as well as critical for businesses. Banks have globally started to utilize the data to obtain useful actionable knowledge via various ways such as sentiment analysis, reputational risk management, product cross-selling, financial crime management, and so on. At present, the banks can easily get benefited from big data where they can quickly extract high-quality information and then convert them into actionable knowledge which can improve the bank's performance [12].

Now-a-days, due to the advancements in technology the credit card is not used as in the past where people now are using e-wallets instead of the physical cards [13]. It was also mentioned that thousands of customers declared bankrupt due to the less controlled purchasing behaviour of the customers. However, some of the credit card facilities are considered as non-secured comparing to other types of debts so it is would be difficult for the banks to collect the debts if customers failed to pay the amount of debt and declared bankrupt [14].

In recent years, the amount of information is increasing due to the growing number of social media users at a rapid speed. Many businesses nowadays have no option other than analyzing the customer's opinions and expressions regarding the products that they experienced. Furthermore, banks must know their customer's opinions on their products to formulate better marketing strategies to enjoy the competitive advantages among the imperfect competition prevailing in the market. Computerized sentiment analysis procedures provide more useful insights on the customer thoughts as well as a satisfaction to make effective actionable decisions. The sentiment analysis will help the banks to take immediate

action to improve its services to the customers especially in the field of debit cards, credit cards and other online services [15].

E. Sentiment Analysis - Techniques and Algorithms

Sentiment analysis is performed in various domains to find out the opinions of the customer for better decision making. Supervised and Unsupervised Machine learning techniques are deployed in classifying and/or predicting the sentiments as positive or negative for the given opinion/reviews [16]. Many researchers have performed sentiment analysis to analyze the movie reviews to classify the sentiments into positive or negative using three different algorithms namely Maximum Entropy classifier, Support Vector Machine (SVM) and Naïve Bayes classifier [17]. Besides that, the three models were augmented using n-grams and the results have shown that the SVM model has outperformed the other two models.

Furthermore, a method based on Artificial Neural Network (ANN) was proposed to classify huge tweets dataset through Hadoop into positive or negative with the suggestion of fuzzy tone where the results were measured according to the accuracy and speed [18]. The results showed that the proposed model was very efficient in dealing with big sentiments datasets better than the small datasets. Also, the results showed that ANN has outperformed SVM and Hidden Markov Model (HMM).

Moreover, other researchers have suggested a new method for the classification of sentiments in the blogospheres and this proposed method was done through the combination of the advantages of Back-Propagation Neural Networks (BPN) and (SO) indexes. The results showed that the suggested method delivered more accurate outcomes and it was noticed that the classifying accuracy was improved with lesser training time [19].

Various libraries were utilized by researchers to perform sentiment analysis and Natural Language Toolkit (NLTK) and Textblob were the most famous. NLTK is considered as a platform that is utilized to build programs in Python and it works with the Natural Language data with statistical NLP application. Also, it includes text processing libraries for classification, tagging, parsing, and so on. Whereas Textblob is a library in Python that is structured on top of NLTK and utilized for NLP. Also, it is considered easier than NLTK as it possesses a simple API which is likely the easiest path to start with the sentiment analysis as well as various text analytics using Python [17].

F. Predictive Modelling in the Credit Card Business

So far, in the credit card market, all the researchers were focusing on credit risk as to the main problem as it is rising every day and has been a long-term issue for the banks. Therefore, various researchers used machine learning algorithms to build models with high prediction accuracy to predict the default rate of credit cardholders. Yet, there are not many pieces of research focusing mainly on the application of machine learning algorithms in predicting the credit card customer satisfaction as indeed a needed one for the current banking industry situation.

A study on credit card fraud detection using machine learning algorithms such as Logistic Regression, Random Forest, Naïve Bayes, and Multilayer Perceptron on the credit card fraud data with class balancing using SMOTE technique. The Random forest obtained the highest accuracy by 99.96% followed by MLP, NB, and LR by 99.93%, 99.23%, and 97.46%, respectively [20].

Another study was conducted on the credit card users default payment using six various machine learning models which are Regression Tree, Nearest Neighbors, SVM Regression, Random Forest Regression, Linear Regression and AdaBoost with the accuracies of 83%, 82%, 85%, 70%, 80%, and 88% respectively where the linear regression got the highest accuracy rate by 88% while the random forest got the lowest accuracy rate by 70% [21].

In the same domain of the credit default risk, six different machine learning models were built namely Naive Bayes, Bayesian networks, J48, random forest, multilayer perceptron, and logistic regression and evaluated using the accuracy, recall, and precision. The logistic regression obtained the highest accuracy (83.108%) whereas Naive Bayes, Bayesian networks, J48, Random forest, and Multilayer perceptron got 82.532%, 82.528%, 82.470%, 82.110%, 81.41%, respectively [22].

The prediction of the default of the credit card users was researched while having implemented balancing techniques such as SMOTE and ADASYN to balance the data [23]. It was noticed that the implemented models (SVM, KNN, Decision Tree and Random forest) achieved the highest accuracies with the normal dataset without any balancing techniques applied except the SVM model that showed the same performance in both normal and balanced data with ADASYN technique. SVM model accuracy showed a very small increase of 0.0025% after balancing the data with SMOTE technique which was considered not an improvement. The model's accuracies were 77.73%, 75.08%, 72.60%, and 80.88% respectively, and noted that the Random forest model achieved the highest accuracy rate of 80.88% with the normal dataset. However, the Recall and ROC values in both balancing techniques ADASYN and SMOTE of all the models were increased and showed values higher than the normal data.

III. MATERIALS AND METHODS

A. Data Collection

The questionnaire is considered as a standard survey tool to collect data in quantitative research. A well-structured questionnaire was used to collect the required data for the research. Since the study population was already known (Young Professionals in Malaysia), the design of the questionnaire was straight-forward and easy with all related instructions. Besides, the questionnaire included both open- and closed-ended questions to minimize the bias and to increase the respondent's rate.

B. Data Analysis

R studio and Tableau were chosen to perform the analytics to attain the specified objectives of this study. Feature

selection is defined as a pre-processing strategy that can be effective in the preparation of data for the various problems in machine learning and data mining. Therefore, the regression and the Correlation-based Feature Selection (CFS) methods were used to choose the most influential attributes. This strategy found useful in acquiring clearer and simple models, with enhanced performance of data mining [24].

The data was pre-processed via various steps such as transformation, cleaning the missing values and detecting outliers if they exist. After cleaning the dataset, descriptive analytics was conducted in which suitable visualizations were used to understand the variables and their relations which revealed some useful insights regarding the scenario. Besides that, after understanding the dataset, predictive analytics particularly sentiment analysis was conducted to figure out the different patterns and opinions of customers regarding the credit card business as well as a predictive model was built to predict the credit card customer satisfaction.

C. Balancing Techniques

A predictive model building using machine learning algorithms is significantly affected by imbalance data. The target variable got two classes (YES and NO) where (NO) was the majority. Therefore, to obtain fair results for the minority class while building the predictive model, class balancing techniques were applied, and this led to balance the dataset distribution.

Three various resampling techniques are generally applied in the class balancing such as over-sampling, under-sampling and hybrid. However, the over-sampling technique was applied over the other techniques due to the less number of records exist in the data set. The variance among oversampling and under-sampling is shown in Fig. 1.

D. Classification Algorithm - Random Forest (RF)

The Random Forest algorithm is a tree-based model in which it is built through the combination of the predictions of different trees where each tree is trained individually [26]. RF model is considered as one of the important competitors to the latest algorithms such as SVM and Boosting. Besides, what differentiates the RF model from the other models is that it is simple as well as fast to apply. Also, it is known for its prominent performance where it gives high accurate predictions. The random forest algorithm can deal with a large number of variables. With each decision split in the RF model, the variables are selected randomly and as a result, this drives the correlation among the trees to decrease [27]. Thus, the power of prediction will be enhanced and as a result, it ends up with higher model efficiency.

Additionally, the RF can assist in extracting the most effective variables that have an impact on the target variables. As a consequence of that, applying this model for predicting the credit card customer satisfaction is substantial. However, the explanation of the RF model is difficult when contrasted to a single decision tree. Therefore, to make the RF model more understandable, there are some possible ways one of them is the estimation of the importance of the features.

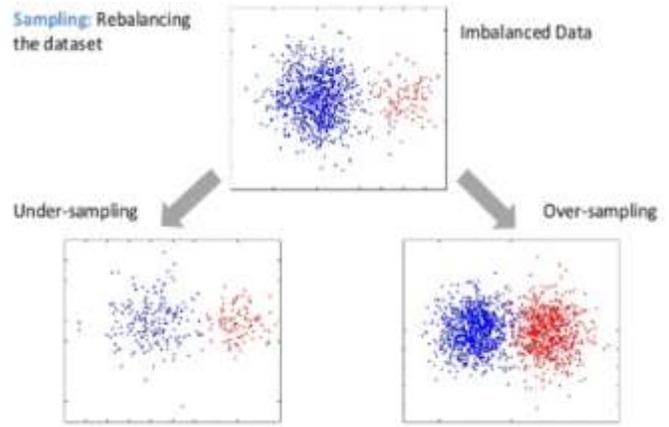


Fig. 1. Over-Sampling and Under-Sampling Methods [25].

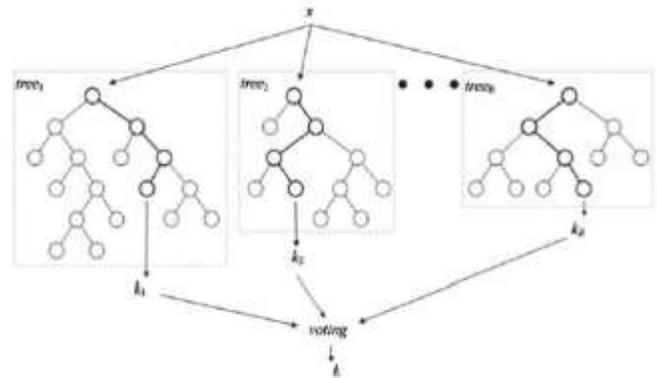


Fig. 2. Random Forest Trees [28].

A tuning parameter named “mtry” was utilized to optimize the RF model performance in which its task falls short in presenting the variables number sampled randomly as candidates within every division of the decision tree as depicted in Fig. 2 [28].

E. Conceptual Framework

As shown in Fig. 3, a conceptual framework is considered as a tool to determine the potential relationships among the independent and dependent variables. Also, it is an incorporated technique to consider the problem of the research with the most suitable approach to the study [29].

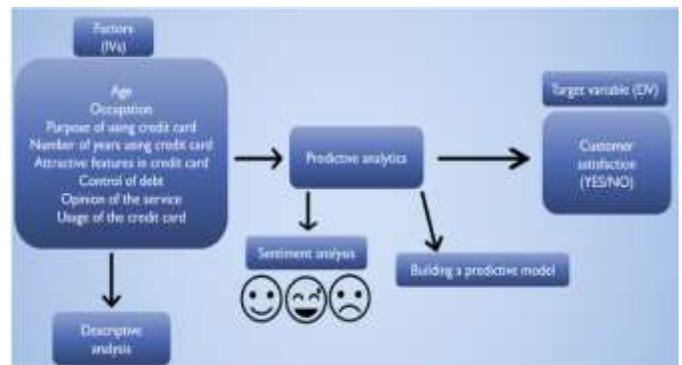


Fig. 3. Conceptual Framework.

IV. DESCRIPTIVE ANALYTICS

The main focus was on one of the most important variables which impact on the behaviour of the customers when using their credit cards. The usage of the credit card variables got five classes in which the promotion class has been selected by 283 individuals. As depicted in Fig. 4, 163 males and 120 females were found who spend using their credit cards under the promotions. Besides, 187 were employed where the remaining 96 were not employed from the 283 individuals. When it comes to the education of this specific group, it was seen that 32 diploma, 22 secondary, 152 were holding a bachelor's degree, 63 masters' degree, and 14 doctorates.

Also, it was clear that the age range 18-24 and 25-34 were representing the majority of this group while the minority was represented by the age range 35 and above. In terms of the satisfaction of the customer, it was seen that from the 283 individuals who were spending with their credit cards during promotions, a total of 222 were not satisfied with the service while only 61 of them were satisfied. Therefore, it could be concluded that the previously analyzed variables were connected and having an impact on the target variable.

So, it could be concluded that young people who were between the age of 18-24 and 25-34 got influenced more than elder people by promotion which drives them to spend more recklessly without thinking about debt problems. Also, the education level got an impact as most of them were bachelor's degree holders with less information about credit card facilities and led them to be dissatisfied.



Fig. 4. Data Visualisation 1.



Fig. 5. Data Visualisation 2.

Fig. 5 mainly focuses on the control of the customers on their credit card debts. It was noticed that a large portion of the respondents was not sure of whether they can control their credit card debts or not. Also, it can be seen that the majority of those people who were not sure belong to the groups that have been using the credit card for 1-2 years (115 individuals) and the group of 3-5 years (84 individual) while the minority of 26 have been using it for 6 years and above. In terms of their education, the bachelors and masters got the majority as 129 and 43 individuals, respectively. Besides, their general opinion about the credit cards with benefits for them represented by 163 individuals while remaining selected no benefits and very beneficial for me by 21 and 41 respectively. When it comes to the target variable, from the 225 individuals 177 were not satisfied with the service compared to 48 who were satisfied. It can be concluded that customers who were using credit cards for short periods mostly holding bachelors' degree and were using it for getting some benefits by spending without thinking about future problems.

V. PREDICTIVE ANALYTICS

A. Modelling: Random Forest Model (RF)

Random Forest is a machine learning algorithm which creates several classification trees through utilizing the bootstrap sampling technique while training the model and, the classification trees produce the final prediction in the test phase. This algorithm was applied to both the unbalanced and balanced dataset and it gave a better performance on the unbalanced dataset.

Fig. 6 displays the Out-of-bag (OOB) error plot which is utilized to measure the prediction error rate of the Random Forest model as well as it identifies the number of trees utilized during the model building process. Also, the error rate used to reduce with the increase in the number of trees. From the three lines, the black line shows the overall error rate, while the green and red lines are referring to one class error, respectively.

1) *Imbalanced dataset*: It can be seen that the increased rate of error decreased when it became stable in all three lines, which means that after 200-300 trees, the error rate did not decrease any more.

Fig. 7 represents the RF model performance with the accuracy of 85.82%.

2) *Balanced dataset (oversampling technique)*: Fig. 8 shows the balanced data using the oversampling technique where both classes (Yes and No) got balanced with 50% each and the RF model obtained 82.83% accuracy.

3) *Random forest model optimization*: The RF model was tuned using the hyperparameters among which the mtry was very prominent. The task of this (mtry) is to represent the features number randomly within every node of the RF tree. Fig. 9 shows that when the mtry is 6 the model got the lowest error rate of 12.62% compared to the mtry 12 and mtry 3 with an error rate of 12.93% and 13.56% respectively. Thus, mtry 6

was selected as the lowest error rate which gave an accuracy of 87.38%.

Random Forest algorithm list out the most important variables while building the classification model and this is considered as one of the most significant outcomes of this algorithm as shown in Fig. 10.

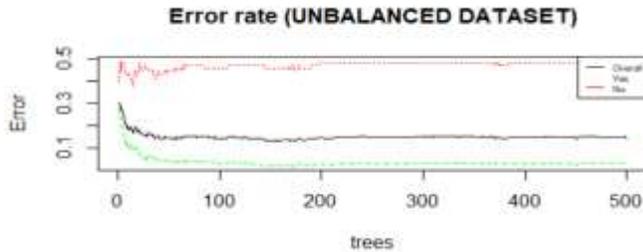


Fig. 6. Unbalanced Dataset Error Rate.

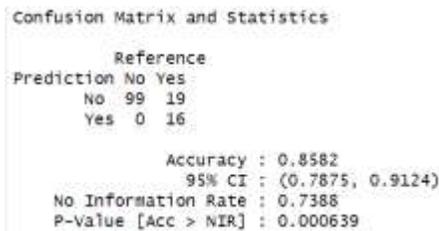


Fig. 7. RF Confusion Matrix for Imbalanced Dataset.

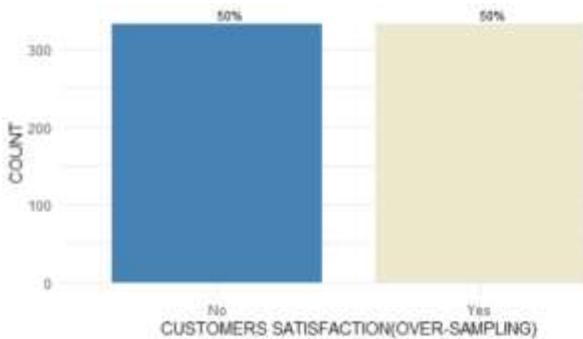


Fig. 8. Target Variable after Balancing.

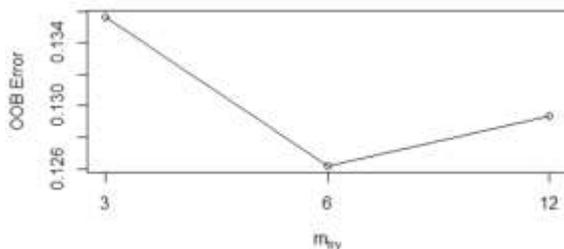


Fig. 9. Error Rate Plot.

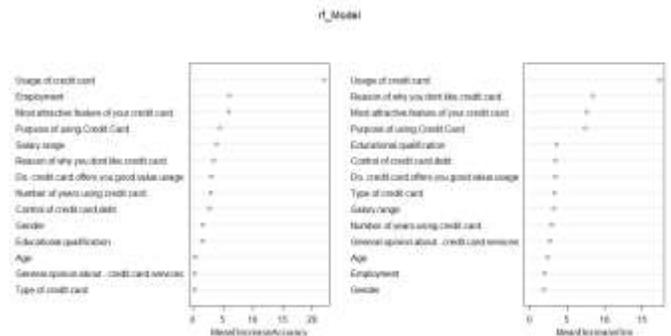


Fig. 10. Factors affecting the Credit Card User's Satisfaction.

The significance of the variables is arranged from top to bottom in which the top variables are with the most significant. Also, there are two kinds of the mean within the figure above which are mean decrease accuracy and mean decrease Gini where the first type represents the most important variable to the least significant by sorting them from top to bottom and the second type represents the measurements of every variable contribution to the similarity of leaves and nodes in the RF model. Besides, it was noted that the usage of the credit card variable got the most effect on the credit card customer satisfaction, while Types of the credit card variable got the least effect.

4) *RF Model experiments*: Table I and Fig. 11 shows all the information's regarding the RF model experiments which state the model accuracy, error rate, recall, and precision percentages.

It could be seen that the optimized model through mtry hyper-parameter achieved the highest accuracy rate of 87.38% followed by the normal dataset with 85.82% and the lowest accuracy rate obtained for the balanced dataset of 82.83%. It is known that the higher recall and precision rates indicate better performance of the model. The precision value for the normal dataset was 100%, where the balanced dataset and the optimized model got 82.75% each. Furthermore, the recall rate for the normal data was 83.89% while it was 82.75% each for the balanced dataset and optimized model. Lastly, the optimized model achieved the lowest error rate of 12.62% followed by the normal data with 14.18% and the highest error rate of 17.17% was achieved for the balanced dataset.

TABLE I. RANDOM FOREST MODEL COMPARISON

Data	Fine-tuning	Accuracy (%)	Recall (%)	Precision (%)
Normal Data(Imbalanced)	-	85.82	83.89	100
Balanced Data(Over-sampling)	-	82.83	82.75	96.96
Optimized Model with Normal Data	Mtry	87.38	82.75	96.96

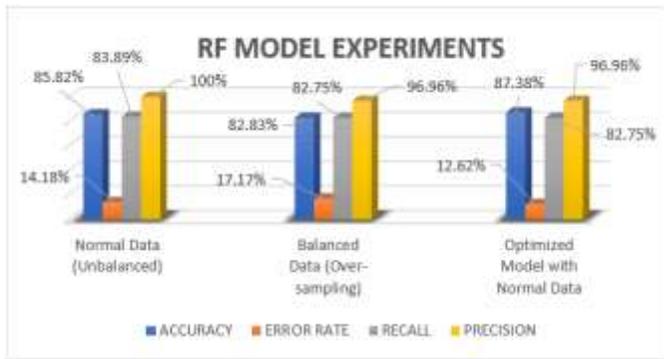


Fig. 11. RF Model Performance.

B. Sentiment Analysis

The Specific opinion of the customer about the credit card variable was used to perform the sentiment analysis using various opinions of the customers. The sentiments of the customers were predicted as positive, neutral, and negative which were represented by 1, 0, and -1, respectively. Fig. 12 shows the predicted sentiments according to the QDAP dictionary and it was presented by 1, 0, and -1 with decision according to the unsupervised rule-based prediction method.

Fig. 13 displays the sentiment plot where it was noticed that negative sentiments were dominant over positive and neutral sentiments. Also, it was noticed from the opinions of the customers that even if the sentence is positive there was negative comment within the sentence itself, for example, "I like the credit card service and it helps me in some situations, however, I will always be worried from fraud and overspending". Almost more than half of the data was in this form where it starts with positive words then stating the negative opinion which caused dissatisfaction with the credit card service. This was the main reason to get more negative compared to positive and neutral sentiments.

The overall sentiments were represented using a pie chart as shown in Fig. 14. The negative sentiments got the highest percentage of 45% followed by positive sentiment (30%), and the least 25% for neutral sentiments.

```
> sentiment$SentimentQDAP
 [1] 0.0000000 0.14285714 0.00000000 0.00000000 0.00000000 0.20000000
 [7] -1.00000000 -1.00000000 -1.00000000 0.50000000 -1.00000000 0.50000000
 [13] 1.00000000 0.16666667 -0.12500000 0.25000000 -0.25000000 0.22222222
 [19] 0.00000000 1.00000000 0.33333333 0.25000000 0.18181818 0.66666667
 [25] 0.27272727 1.00000000 0.14285714 0.20000000 0.00000000 1.00000000
 [31] -1.00000000 0.50000000 -0.16666667 1.00000000 0.16666667 0.00000000
 [37] -0.16666667 0.00000000 0.00000000 0.16666667 0.23076923 1.00000000
 [43] 0.40000000 0.09090909 1.00000000 0.25000000 0.00000000 1.00000000
 [49] -1.00000000 0.00000000 -0.04347826 -0.33333333 0.27777778 0.66666667
 [55] 1.00000000 0.00000000 0.06896552 0.00000000 0.13636364 0.35294118
 [61] 0.00000000 0.20000000 -0.07692308 0.00000000 0.00000000 0.25000000

> convertToDirection(sentiment$SentimentQDAP)
 [1] neutral positive neutral neutral neutral neutral positive negative negative
 [9] negative positive negative positive positive positive negative positive
 [17] negative positive neutral positive positive positive positive positive
 [25] positive positive positive positive neutral positive negative positive
 [33] negative positive positive neutral positive neutral neutral positive
 [41] positive positive positive positive positive positive positive neutral positive
 [49] negative neutral negative negative positive positive positive neutral
 [57] positive neutral positive positive neutral positive negative neutral
```

Fig. 12. Sentiment Classification.

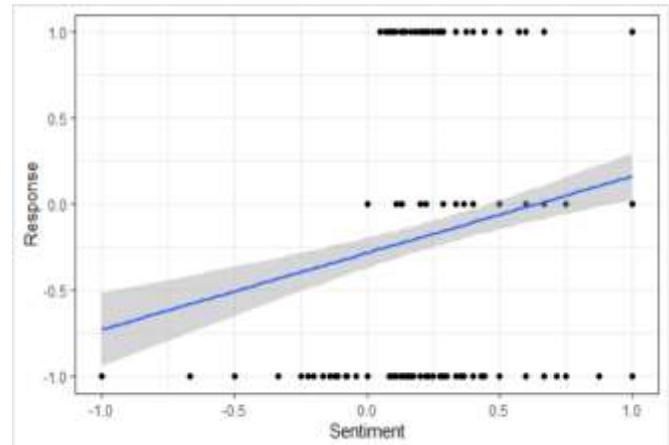


Fig. 13. Sentiment Analysis Plot.

SENTIMENTS COMPARISON

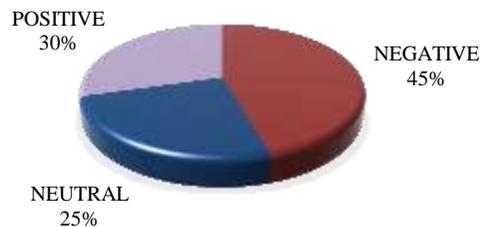


Fig. 14. Sentiments Comparison.

According to the results obtained from the sentiment analysis, it can be seen that most of the users were dissatisfied and had negative opinions about the credit card services which may indicate the existence of other problems such as customers who are not able to carry out their responsibilities with their credit card spending habits and it ends up with debts and financial problems. Also, it may indicate that some customers are not financially literate, and they keep spending more money without realizing that they put themselves in difficult situations that drive them to damage their financial future while the banks are creating more profits from the interests charged to the credit cards.

VI. CONCLUSION

The main aim of this study was to perform predictive analytics particularly sentiment analysis as well as building a predictive model to predict customers' satisfaction on the credit card services. The data mining and data visualization were utilized to perform all the tasks as well as to explore all the variables and show the effect on the credit card customer satisfaction. The oversampling technique was applied to perform class balancing and a predictive model was built using the Random Forest algorithm. Also, three different experiments were conducted using normal data, balanced data, and optimized model with normal data to achieve the objectives of this study.

As stated earlier, the best accuracy of 87.38% was obtained for the optimized model through the mtry hyperparameter. On the other hand, sentiment analysis on the

specific opinion of the customers on credit card services was performed and found that 45% of the responses reflected negative sentiments. It was concluded that the people in the age range of 18 to 34 were explicitly utilizing the credit card and generally fail to pay back their debts or always worried of fraud and risk of spending more carelessly. The outcome obtained from this study would be more useful to the decision-makers in the banking sector concerning the development of the credit card business in future.

There can still be improvements in applying robust machine learning algorithms and/or deep learning architecture with proper optimization techniques to build a more effective classification model. Also, socio-economic variables can be included in the analytics process to get a high-quality insight which would be more useful to the decision-makers. However, limited time and restricted access to data due to the current pandemic situation were the major limitations in this study. So, increasing the sample size of the data as well as adding new variables is recommended to obtain better results in all means in future.

REFERENCES

- [1] C. Bertaut, C. and Haliassos, M. "Credit Cards: Facts and Theories," SSRN Electronic Journal, 2005, doi:10.2139/ssrn.931179.
- [2] Agarwal, S. "Why Do Banks Reward Their Customers to Use Their Credit Cards?," 2011, FRB of Chicago Working Paper.
- [3] Chatterjee, P. & Rose, R. "Do Payment Mechanisms Change the Way Consumers Perceive Products," Journal of Consumer Research, 38(6), 2012, pp.1129-1139, doi: 10.1086/661730.
- [4] E. Halim, R., Adiwijaya, K., Haryanto, J. and Firmanzah, F. "The Propensity of Young Consumers to Overspend on Credit Cards: Decomposition Effect in the Theory of Planned Behavior," Journal of Economics, Business and Management, 4(10), 2016, doi: 10.18178/joebm.2016.4.10.459.
- [5] The sun daily. "3.6m credit card holders in Malaysia as at June," [online] p.A single page. Available at: <https://www.thesundaily.my/archive/36m-credit-card-holders-malaysia-june-ITARCH473259>, 2017, [Accessed 4 Jul. 2020].
- [6] UKEssays. "Overview of Credit Cards In Malaysia," Marketing Essay. [online]. Available from: <https://www.ukessays.com/essays/marketing/overview-of-credit-cards-in-malaysia-marketing-essay.php?vref=1>, November 2018, [Accessed 14 December 2020].
- [7] Lim, W. M., Ng, W. K., Chin, J. H. and Boo, A. W. X. "Understanding Young Consumer Perceptions on Credit Card Usage: Implications for Responsible Consumption," Contemporary Management Research, 10(4), 2014, doi:10.7903/cmr.11657.
- [8] Teoh, W. M. Y., Chong, S. C. and Yong, S. M. "Exploring the factors influencing credit card spending behavior among Malaysians," International Journal of Bank Marketing, 31(6), 2013, DOI: 10.1108/IJBM-04-2013-0037.
- [9] Khare, A., Khare, A. and Singh, S. "Factors affecting credit card use in India," Asia Pacific Journal of Marketing and Logistics, 24(2), 2012, pp. 236-256, doi: 10.1108/13555851211218048.
- [10] Hussin, S. R., Kassim, S., and Jamal, N. "Credit Card Holders in Malaysia: Customer Characteristics and Credit Card Usage," Int. Journal of Economics and Management, 7(1), 2013, pp.108 – 122.
- [11] Ahmed, Z. U., Ismail, I., Sadiq Sohail, M., Tabsh, I., and Alias, H. "Malaysian consumers' credit card usage behaviour," Asia Pacific Journal of Marketing and Logistics, 22(4), 2010, pp.528–544. doi:10.1108/13555851011090547.
- [12] Srivastava, U. and Gopalkrishnan, S. "Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks," 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15), 2015, pp.643–652. doi: 10.1016/j.procs.2015.04.098.
- [13] New Straits Times. "Number of bankrupts due to credit card usage relatively low," [online] Available at: <https://www.nst.com.my/business/2018/03/345048/number-bankrupts-due-credit-card-usage-relatively-low>, 2019, [Accessed 12 Jul. 2020].
- [14] Ismail, S. "Watch out for credit card debt. New Straits Times," [online]. Available at: <https://www.nst.com.my/opinion/letters/2019/02/462125/watch-out-credit-card-debt>, 2019, [Accessed 1 Jul. 2020].
- [15] N.Sumathi and T.Sheela. "Opinion Mining Analysis in Banking System using Rough Feature Selection Technique from Social Media Text," International Journal of Mechanical Engineering and Technology (IJMET), 8(12), 2017, pp.274–289.
- [16] Borele, P. and Borikar, D. A. "An Approach to Sentiment Analysis using Artificial Neural Network with Comparative Analysis of Different Techniques," Journal of Computer Engineering (IOSR-JCE), 18(2), 2016, pp.64-69.
- [17] Pang, B., Lee, L. and Vaithyanathan, S. "Thumbs up? Sentiment Classification using Machine Learning Techniques," 2002, Philadelphia, s.n.
- [18] Jian, Z., Chen, X. and Han-shi, W. "Sentiment classification using the theory of ANNs," The Journal of China Universities of Posts and Telecommunications, Volume 17, 2010, pp.58-62.
- [19] Chena, L. S., Liub, C. H. and Chiu, H.-J. "A neural network based approach for sentiment classification in the blogosphere," Journal of Informetrics, 5(2), 2011, pp.313-322.
- [20] Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., and Anderla, A. "Credit Card Fraud Detection - Machine Learning methods," INFOTEH-JAHORINA, 2019, 18th International Symposium.
- [21] Ullah, M. A., Alam, M. M., Sultana, S. and Toma, R. S. "Predicting Default Payment of Credit Card Users: Applying Data Mining Techniques 2018," 2nd Int. Conf. on Innovations in Science, Engineering and Technology (ICISSET), 2018, Chittagong, Bangladesh, s.n.
- [22] Çırsar, B. and Unal, D. "Comparison of Data Mining Classification Algorithms Determining the Default Risk," Scientific Programming, 2019, pp.1-8.
- [23] Shetty, A. S. and R, M. "Prediction Of Default Credit Card Users Using Data Mining Techniques," International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(7), 2019, pp.2278-3075.
- [24] Jundong, L., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J. and Liu, H. "Feature Selection: A Data Perspective," ACM Computing Surveys (CSUR), 50(6), 2016, pp.1–44. doi: 10.1145/3136625.
- [25] Walimbe, R. "Handling Imbalanced Dataset in Supervised Learning Using Family of SMOTE Algorithm," Data Science Central. Available at: <https://www.datasciencecentral.com/profiles/blogs/handling-imbalanced-data-sets-in-supervised-learning-using-family>, 2017, (Accessed: 31 August 2019).
- [26] Denil, M., Matheson, D. and Freitas, N. De. 2014. Narrowing the Gap : Random Forests In Theory and In Practice, in Proceedings of the 31st International Conference on Machine Learning. Beijing, China: JMLR, pp. 1–9. Available at: <http://proceedings.mlr.press/v32/denil14.pdf>.
- [27] Biau, G. "Analysis of a Random Forests Model," Journal of Machine Learning Research, 13(1), 2012, pp.1063–1095. Available at: <http://www.jmlr.org/papers/volume13/biau12a/biau12a.pdf>.
- [28] Esteves, G. and Mendes-Moreira, J. "Churn Prediction in the Telecom Business," Eleventh International Conference on Digital Information Management (ICDIM), Porto, Portugal: IEEE, 2016, pp.254–259. doi: 10.1109/ICDIM.2016.7829775.
- [29] Liehr, P., Smith M. J. "Middle range theory: Spinning research and practice to create knowledge for the new millennium," Advances in Nursing Science, 21(4), 199, pp.81-91.